



# Intonation tutor by SPIRE (In-SPIRE): An online tool for an automatic feedback to the second language learners in learning intonation

Anand P A<sup>1</sup>, Chiranjeevi Yarra<sup>2,a</sup>, Kausthubha N K<sup>2</sup>, Prasanta Kumar Ghosh<sup>2,b</sup>

<sup>1</sup>National Institute of Technology Karnataka (NITK), Surathkal-575025, India

<sup>2</sup>Electrical Engineering, Indian Institute of Science (IISc), Bangalore-560012, India

{<sup>a</sup>chiranjeevi, <sup>b</sup>prasantg}@iisc.ac.in

## Abstract

In spoken communication, intonation often conveys meaning of an utterance. Thus, incorrect intonation, typically made by second language (L2) learners, could result in miscommunication. We demonstrate In-SPIRE tool that helps the L2 learners to learn intonation in a self-learning manner. For this, we design an interactive self-explanatory front end, which is also used to send learner's audio and hand-shake signals to the back-end. At the back-end, we implement a system that takes the learner's audio against a specific stimuli and computes pitch patterns representing the intonation. For this, we apply pitch stylization on each syllable segment in the audio. Further, we compute a quality score using the learner's patterns and the respective ground-truth patterns. Finally, the score, the patterns of the learners and the ground-truth are sent to the front-end for display as a feedback to the learners. Thus, the learner could correct any mismatch in his/her intonation with respect to the ground-truth. The proposed tool benefits the learners who do not have access to effective spoken language training.

## 1. Introduction

Intonation often adds meaning to words and word group in spoken English communication [1]. Hence, incorrect intonation would result in miscommunication. Thus, in the second language (L2) training, L2 learners require to learn the intonation for better communication. Further, the L2 learners could be benefited from an automated system that evaluates the learner's proficiency in intonation [2, 3]. Accessing such systems via online interface from the remote locations could further benefit the people for whom effective training methods are not easily accessible. In this demo, we present In-SPIRE<sup>1</sup> tool that assess intonation of a learner's audio with respect to an expert and provides feedback to the learner in an automated way. To the best of our knowledge, there have been no such online tools available.

## 2. Proposed architecture

The architecture of the proposed self-learning tool is shown in Figure 1. It has two major components – front-end (user interface) and back-end (web-server). The front-end is available at the learner's location and the back-end is situated at our location. Both the front-end and back-end communicate via Internet.

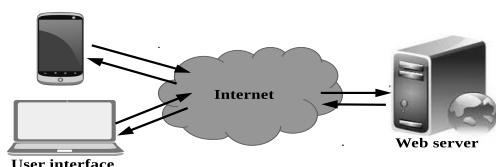


Figure 1: Architecture of the proposed self-learning tool

### 2.1. Front-end

Figure 2a shows the dashboard after the user login. On this dashboard, a stimuli to read is shown along with three buttons

for UI control – 1) submit the recording 2) Previous 3) Next. In addition to these buttons, user can record the voice by clicking the button with microphone symbol and the recorded voice can be listened back using play button. Thus, this allows the user to verify their recording. If the recorded voice is not proper, it can be re-recorded and listened till the user's expected recording is achieved. Following this, on clicking "submit the recording" button, user can view the intonation quality of the recorded stimuli with respect to the intonation of the expert's voice in the window shown in dotted black rectangular box in Figure 2a. In the meantime, the "submit the recording" is replaced with two buttons – 1) Listen to expert 2) Know your performance. In case, the user wants to correct the pronunciation according to the experts, he/she can view the changes in the pitch patterns as well as the pronunciation of the expert's audio using "Listen to expert" button. Figure 2b shows the play button to listen the expert's audio and the respective pitch patterns for each syllable. Using the audio and pitch patterns, learners can adapt themselves according to the experts voice.

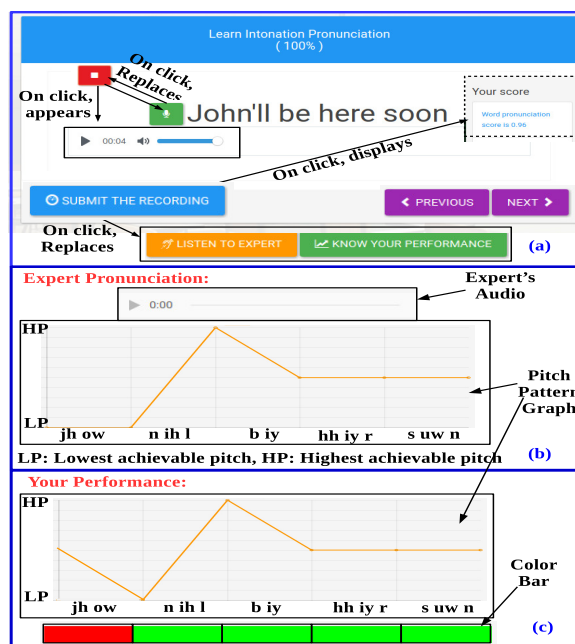


Figure 2: An illustration of the proposed front-end that appears on – a) user login b) clicking "Listen to expert" c) clicking "Know your performance"; considering the exemplary screen shots taken from the proposed In-SPIRE tool.

Moreover, learners can obtain the feedback on their pitch pattern by clicking "Know your performance" button. Figure 2c shows the pitch pattern changes obtained for an exemplary learner's recording. In this figure, patterns are shown for all syllables and the learner's performance at each syllable is also indicated with a color bar. The green color block in the color bar indicates a match between the intonation patterns in the learner's and the expert's pronunciation. Similarly, the red color block

<sup>1</sup><https://spire.ee.iisc.ac.in/In-SPIRE/>

indicates a mismatch. Using this feedback, learners can practice themselves to achieve expert's quality intonation in their pronunciation for the current stimuli. On completion of the practice, learners can choose previous or next stimuli by clicking "Previous" or "Next" buttons respectively.

## 2.2. Back-end

In the intonation training, L2 learners require to vary their pitch according to a sequence of discrete symbols called tones [1] that represent intonation of the expert's audio. In this demo, we consider the tones as the pitch patterns. The tones, in general, are independent of speaker specific variations and there are four different types indicating following four pitch variations – 1) the rise tone: pitch rises from a low value to high value 2) the fall tone: pitch falls from a high value to low value 3) the high tone: pitch value remains fixed at a high value 4) the low tone: pitch value remains fixed at a low value. However, L2 learners might not have prior knowledge of the tones. Thus, in this demo, we construct a graph from the tone sequence to show the required pitch changes directly to L2 learners.

### 2.2.1. Pitch pattern (tone) extraction

We compute the tones for each syllable from the pitch variations within respective segments as follows:

- We approximate the pitch variations in each syllable segment with a straight line by minimizing absolute difference between the approximated line and the pitch variations [4]. Let  $m$  and  $c$  are the slope and abscissa of the line respectively.
- We replace the pitch variations in each syllable with the respective line segments and perform mean and range normalization.
- We consider that the syllable has the rise tone when  $|m| \geq m_t$  and is positive, where  $m_t$  is a threshold value. If  $m$  is negative and  $|m| \geq m_t$  then the fall tone is considered. Similarly, if  $|m| < m_t$  and  $c < 0$  then the low tone is considered otherwise the high tone is considered. In-SPIRE tool uses  $m_t = 1$ , determined empirically.

### 2.2.2. Pitch pattern graph construction

In order to obtain the graph, we consider fixed lower and upper limits across all the users as well as the expert. Further, without loss of generality, we indicate these limits as lowest and highest achievable pitch values by the respective user and the expert on the graph in a qualitative manner. With this representation, we assume that the constructed graph becomes specific to each user. Using these lower and upper limits, we construct the graph as follows:

- We chunk the entire tone sequence into sequence of groups containing sequence of same tones.
- For each group of rise (fall) tones, we construct a line joining the lower (upper) limit at the starting of the syllable belonging to the first tone to the upper (lower) limit at the end of the syllable belonging to the last tone.
- For the group of low (high) tones, we consider the lower (upper) limit for the entire length covered by the syllables in that group.

For the graph construction, we consider the length of all the syllables identical irrespective of their actual durations.

### 2.2.3. Intonation quality score computation:

We compute intonation quality of the L2 learners in comparison with the expert intonation using the graphs constructed from the steps described in Section 2.2.2. From each learner and expert graph, we sample uniformly and obtain sequence

of values representing learners and experts intonation respectively. Following this, we apply dynamic time warping between these sequences and obtain time aligned sequences for the expert and the learner. Further, we compute Pearson correlation co-efficient between these time aligned sequences and consider it as the intonation quality score of the learner.

## 3. Demonstration

For the demo, in order to obtain the syllable transcriptions for the learners audio, first, we perform force-alignment of the audio considering its transcription using Kaldi speech recognition tool kit and estimate phoneme transcriptions and its boundaries. In the force-alignment process, we consider a lexicon containing pronunciations for each word. Next, we perform syllabification on phoneme transcriptions of each word in an utterance separately using P2TK syllabifier [5] and obtain syllable transcriptions and its boundaries. We use JavaScript and HTML for front-end coding and Node.js for back-end coding [6]. We set-up the server using LAMP (Linux, Apache, MySQL, PHP) stack on Ubuntu 14.04 operating system. At the back-end, we use Python programming language to implement tone extraction and graph construction. We consider a set of 80 stimuli taken from the training material used for spoken English training [1]. We use YIN algorithm for pitch computation [7]. In order to obtain expert's audio, we record the stimuli from a voice-over artist, proficient in British English spoken communication.

## 4. Conclusion

We present an online self-learning tool named In-SPIRE that provides an automatic feedback to L2 learners for learning intonation. We design the front-end for the tool with Javascript and HTML coding and the back-end with Node.js coding and Python programming language. Using this tool, we provide a feedback by showing the intonation quality score of the learners with respect to experts as an assessment measure as well as the learner's and the expert's intonation pitch patterns as corrective measures. Further investigations are required to develop better pitch pattern extraction methods.

## 5. Acknowledgement

We would like to thank Chetak R M of Maiyas Enterprises for his constant support in developing In-SPIRE tool. We also thank Pratiksha Trust for their support.

## 6. References

- [1] J. D. O'Connor, *Better English Pronunciation*. Cambridge University Press, 1980.
- [2] J. P. Arias, N. B. Yoma, and H. Vivanco, "Automatic intonation assessment for computer aided language learning," *Speech communication*, vol. 52, no. 3, pp. 254–267, 2010.
- [3] D. Ke and B. Xu, "Chinese intonation assessment using SEV features," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 4853–4856, 2009.
- [4] O. J. Karst, "Linear curve fitting using least deviations," *Journal of the American Statistical Association*, vol. 53, no. 281, pp. 118–132, 1958.
- [5] J. Tauberer, "P2TK automated syllabifier," Available at <https://sourceforge.net/p/p2tk/code/HEAD/tree/python/syllabify/>, last accessed on 10-03-2018.
- [6] M. Cantelon, M. Harter, T. Holowaychuk, and N. Rajlich, *Node.js in Action*. Manning Publications, 2017.
- [7] A. De Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.